BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e137827
doi: 10.3897/biss.8.137827

OPEN ACCESS

Conference Abstract

DoeDat: Enhanced Roundtripping of Crowdsourced Specimen Annotations

Mathias Dillen‡, Maarten Trekels‡

‡ Meise Botanic Garden, Meise, Belgium

Corresponding author: Mathias Dillen (mathias.dillen@plantentuinmeise.be)

Received: 25 Sep 2024 | Published: 25 Sep 2024

Citation: Dillen M, Trekels M (2024) DoeDat: Enhanced Roundtripping of Crowdsourced Specimen Annotations. Biodiversity Information Science and Standards 8: e137827. https://doi.org/10.3897/biss.8.137827

Abstract

The DoeDat platform was launched by Meise Botanic Garden in 2018 to capture label data from imaged herbarium specimens by inviting volunteer contributors (Groom et al. 2018). It has since facilitated data capture from specimens of other natural history collections (Helminger et al. 2020, Mitrache et al. 2023), as well as digitised content from various other disciplines, such as historical photographs, posters and postcards. Volunteers may simply transcribe handwritten and/or typed text, but often also interpret the sparse and scattered information on the image, including trying to georeference its original location. As of April 2024, almost 650.000 tasks have been completed, of which more than 470.000 were herbarium specimens from Meise.

DoeDat supports domain standards, including Darwin Core, and follows most of the currently drafted MIDS (Minimum Information about a Digital Specimen) guidelines as to what data is captured for natural history specimens. However, images have to be pre-loaded into the server storage for each project and captured data gets exported as one or more CSV files per project. These data files then still need to be processed before they can be ingested into the local management system (Engledow et al. 2023). Often the data are also subjected to additional quality control before they get openly published. This can result in the pipeline from image to openly published annotations being quite time and labour-consuming.

As the biodiversity infrastructure landscape moves more towards FAIR (Findable, Accessible, Interoperable, Reusable) open data, DoeDat will adapt accordingly. This

© Dillen M, Trekels M. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

includes digital objects that are easy to annotate. Furthermore, image servers following IIIF (International Image Interoperability Framework) greatly standardise the access and portability of media content, drastically changing the way images are being dealt with. We envision upgrading the DoeDat platform to load images and any required metadata as IIIF manifests, greatly streamlining the process of adding new content and tracking provenance. The transcriptions should be accessible for external systems, loading the updated image manifests and publishing them as annotations such as nanopublications.

Keywords

label transcription, IIIF, MIDS, FAIR, data provenance

Presenting author

Mathias Dillen

Funding program

DoeDat received funding through the Flemish Government's DOE! and DOE!2 projects. It also received funding from FWO through the DiSSCo Flanders project and from the Flemish Open Science Board. MD received an FWO travel grant to attend the TDWG 2024 conference.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Engledow H, Bogaerts A, De Smedt S, Tack W (2023) Planning the Migration to a New Database: Implications for the Collections of Meise Botanic Garden. Biodiversity Information Science and Standards 7 https://doi.org/10.3897/biss.7.111422
- Groom Q, De Smedt S, Veríssimo Pereira N, Bogaerts A, Engledow H (2018) DoeDat, the Crowdsourcing Platform of Meise Botanic Garden. Biodiversity Information Science and Standards 2 https://doi.org/10.3897/biss.2.26803
- Helminger T, Weber O, Braun P (2020) Digitisation of the LUX herbarium collection of the National Museum of Natural History Luxembourg. Bulletin de la Société des naturalistes luxembourgeois 122: 147-152. [In English]. URL: https://www.researchgate.net/publication/342957093

- Mitrache L, Paleco C, Huyse T (2023) Mapping and SWOT analysis of Citizen Science actions at the Royal Belgian Institute of Natural Sciences and the Royal Museum for Central Africa. RBINS/RMCA. Last accessed 2024-07-24. URL: https://www.africamuseum.be/publication_docs/Mapping%20-%20SWOT%20-%20CS%20-%20RBINS%20-%20RMCA%20-2023.pdf